

A New Echo Hiding Algorithm with High Robustness

Mo Zhou¹, Shun-Dong Li², Xiang-Yang Luo^{3, 4}, Dao-Shun Wang¹.

¹Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

²School of Computer Science, Shaanxi Normal University, Xi'an 710062, China

³China Institute of Electronic Equipment System Engineering, Beijing 100141, China

⁴Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China

daoshun@mail.tsinghua.edu.cn (Dao-Shun Wang). Tel.:+86-10-62782930.

Abstract

Echo hiding method boasts high watermark capacity but not high robustness. We propose a new echo hiding method taking advantage of both channels to replace the original single echo hiding algorithm. Then, we propose a new echo hiding algorithm combined with spread spectrum algorithm in Modulated Complex Lapped Transform domain. The experimental results show higher robustness compared with original echo hiding method.

Keywords

Echo Hiding; Robustness; Channel; Capacity

Introduction

Echo hiding method was described for the first time in [1]. An echo can be considered as a delayed version of the signal itself. The delay can be made small so that the echo is not audible. Many algorithms based on echo hiding have been developed, such as single echo, double echo, forward backward echo and time spread echo hiding [2, 3, 4, 5] and their improvements. Yousof Erfani et al. [6] proposed three methods based on single and double echo hiding. Foo Say Wei et al. [7] embedded watermark bit 0 and 1 into different channels to achieve higher robustness. Duan et al. [8] introduced a developed cepstrum method which improved the accuracy by making use of the cepstral value of the original signal. Experimental results in [8] show that Echo hiding method has high watermark capacity without high robustness.

The Modulated Lapped transform (MLT) is commonly used to implement block transform coding in video and audio compression. It allows for perfect reconstruction, has almost optimal performance for transform coding of a wide variety of signals with no blocking artifacts [9]. In [10] the modulated complex lapped transform (MCLT) was given as a simple

extension to the MLT and preserved the advantages of it. After that, some research has been done to improve the speed of MCLT algorithm [11-15]. We use MCLT to improve the robustness of echo hiding method in [1]

The outline of the paper is as follows. We propose a new echo hiding method for stereo signals in Section 2 and combine MCLT with our echo hiding method to improve robustness. The evaluation results and discussions are presented in Section 3. Finally the conclusions are given in Section 4.

Capacity Enlargement—Combining Echo Hiding with Spread Spectrum

Echo hiding method boasts a relatively large data capacity. The spread spectrum method embeds watermarks in the frequency domain while the echo hiding method works in the time domain. What's more, both methods process audio sequences block by block. Based on [1], we propose a new stereo echo hiding method below. Then, we try to combine it with the spread spectrum method in MCLT domain to improve robustness.

Our Improved Echo Hiding Method

We propose a new echo hiding scheme for stereo signals. Both channels are used to encode a watermark bit with different delay values. To encode watermark bit 0, echoes with delay value d_0 are embedded in the left channel and echoes with delay value d_1 are embedded in the right channel. To encode watermark bit 1, echoes with delay value d_1 are embedded in the left channel and echoes with delay value d_0 are embedded in the right channel.

When extracting the embedded watermark, we compare the difference of cepstrum value in two delay points of both channels.

The embedding procedure in our method based on [1]

is as follows .

Input: an stereo audio signal portion $x_i(n)$, echo bit e , initial echo amplitude a , n varies from 0 to $M-1$, $e \in \{0,1\}$, M is the number of samples in the block
Watermark embedding procedure: 1) Divide the stereo audio signal portion $x_i(n)$ into left channel portion $x_i^l(p)$ and $x_i^r(q)$, p and q vary from 0 to $\frac{M}{2}-1$ 2) Embed the echo bit into both channels of the block and get the resulted sequence(see [1]) $y_i^l(p) = x_i^l(p) + a * x_i^l(p - d_e)$ $y_i^r(q) = x_i^r(q) + a * x_i^r(q - d_{1-e})$ 3) Recombine the left portion and the right portion and get the watermarked audio portion $y_i(n)$
Output: an processed audio signal block $y_i(n)$

The extracting procedure in our method is as follows.

Input: an processed audio signal block $y_i(n)$
Watermark extracting procedure: 1) Divide the stereo audio signal portion $y_i(n)$ into left channel portion $y_i^l(p)$ and $y_i^r(q)$, p and q vary from 0 to $\frac{M}{2}-1$ 2) Compute the cepstrum of both channels(see [1]) $c_l(p) = F^{-1}(\log F(y_i^l(p)))$ $c_r(q) = F^{-1}(\log F(y_i^r(q)))$ 3) Decide the echo hiding bit e $e = \begin{cases} 0 & \text{if } c_l(d_0) - c_l(d_1) > c_r(d_0) - c_r(d_1) \\ 1 & \text{if } c_l(d_0) - c_l(d_1) < c_r(d_0) - c_r(d_1) \end{cases}$
Output: extracted echo bit e

TABLE 1 CORRECT BIT RATE (CBR) AND SIGNAL-TO-NOISE-RATIO (SNR) OF SINGLE ECHO HIDING METHOD AND OUR PROPOSED ECHO HIDING METHOD

audio sequence	single echo hiding[1]		our proposed echo hiding	
	CBR	SNR	CBR	SNR
clip 1-10	76.09%	16.008	79.53%	17.693
clip 11-20	75.78%	17.540	78.13%	18.784
clip 21-30	76.72%	18.653	78.59%	20.118
clip 31-40	75.16%	20.927	76.09%	22.147
clip 41-50	74.22%	17.406	81.09%	18.749
clip 51-60	80.16%	15.564	85.78%	17.249
clip 61-70	84.69%	20.686	84.84%	22.019
average	77.54%	18.112	80.58%	19.537

TABLE 2. CBR OF SINGLE ECHO HIDING METHOD AND OUR PROPOSED ECHO HIDING METHOD UNDER VARIOUS ATTACKS

Attacks	single echo hiding[1]	our proposed echo hiding
Re-sampling(22.05kHz)	66.75%	82.83%
MP3 compression	56.50%	56.47%
Noise attack	62.70%	65.11%

Low-pass filtering	61.56%	73.37%
High-pass filtering	63.39%	64.15%

We use Sound Quality Assessment Material (SQAM) audio[16] as test material to compare our echo hiding method with the original single echo hiding.

Table 1 shows that our proposed method has higher extracting rate and better conceptual transparency compared to the single echo hiding method. Table 2 shows that our proposed echo hiding method is more robust than single echo hiding under various attacks except MP3 compression.

It is necessary to point out that the comparison between single echo hiding and our proposed echo hiding are based on the same embedding capacity of one watermark bit in 2048 samples of two channels, the same decay rate of 0.5 and initial amplitude of 0.1 and 0.08 respectively. Note that our method is computationally more complex than the method in [1].

Watermark Embedding Procedures of the Proposed Approach

Next we will combine the echo hiding method

proposed above the spread spectrum method in MCLT domain and get a new scheme that works both in the time domain and the transform domain. In the encoder side, we've tried two combining orders and find out that echoes embedded after the MCLT transform will result in better performance in robustness. In the decoder side, since both methods just read the audio sequence rather than modify it, the order doesn't matter.

The embedding procedure is as follows with step 1-3 comes from [17].

Watermark Extracting Procedures of the Proposed Approach

The extracting procedure is as follows with step 1-3 comes from [17].

Let $p \cdot q$ denote the normalized inner product of vector p and q , i.e., $p \cdot q \equiv N^{-1} \sum_i p_i q_i$

Input: audio signal blocks $x_i(k)$, magnitude change a , permuted watermark chip $chip_s(k)$, k varies from 0 to $M-1$, s varies from 0 to $c-1$, c is the number of all possible watermark characters, echo bit e , $e \in \{0,1\}$, M is the number of samples in the block

Watermark embedding procedure:

- 1) Compute the analysis window $h_a(n)$ (see [10])
- 2) Perform MCLT transform on $x_i(k)$ and get MCLT coefficients

$$X_i(k) = \sum_{n=0}^{M-1} x_{i-1}(n)p_a(n,k) + \sum_{n=M}^{2M-1} x_i(n-M)p_a(n,k)$$
- 3) Modify $X_i(k)$ according to the corresponding watermark bit $chip(k)$ and get

$$X'_i(k) = \begin{cases} X_i(k) \cdot a & \text{if } chip_s(k) = 1 \\ X_i(k) \cdot (1/a) & \text{if } chip_s(k) = 0 \end{cases}$$
- 4) Perform inverse MCLT transform on $X'_i(k)$ and get $x'_i(k)$

$$x'_i(k) = \sum_{n=0}^{M-1} X'_i(n)p_s(k,n)$$
- 5) Divide the stereo audio signal portion $x'_i(k)$ into left channel portion $x'_i(p)$ and
 $x'_i(q)$, p and q varies from 0 to $\frac{M}{2} - 1$
- 6) Embed the echoes into both channels of the block (see [1]) and get the resulted sequence

$$y_i^l(p) = x_i^l(p) + a * x_i^l(p - d_e)$$

$$y_i^r(q) = x_i^r(q) + a * x_i^r(q - d_{1-e})$$
- 7) Recombine the left portion and the right portion and get the watermarked audio portion $y_i(k)$

Output: processed audio signal blocks $y_i(k)$

Input: watermarked audio signal blocks $y_i(k)$, k varies from 0 to $M-1$, M is the number of samples in the block

Watermark extracting procedure:

- 1) Compute the analysis window $h_s(n)$ (see [10])
- 2) Perform MCLT transform on $y_i(k)$ and get MCLT coefficients

$$Y_i(k) = \sum_{n=0}^{M-1} y_{i-1}(n)p_s(n,k) + \sum_{n=M}^{2M-1} y_i(n-M)p_s(n,k)$$
- 3) Compute the correlations of $Y_i(k)$ with all possible watermark chips in the pool and get the extracted watermark chips $chip_s(k)$

$$Correlation(Y_i, chip_s) = Y_i \cdot chip_s = \text{Max}\{Y_i \cdot chip_t\} \text{ for all possible } t$$
- 4) Divide the stereo audio signal portion $y_i(k)$ into left channel portion $y_i^l(p)$ and $y_i^l(q)$, p and q varies from 0 to $\frac{M}{2} - 1$
- 5) Compute the cepstrum of both channels (see [1])

$$c_l(p) = F^{-1}(\log F(y_i^l(p)))$$

$$c_r(q) = F^{-1}(\log F(y_i^r(q)))$$
- 6) Decide the echo hiding bit e

$$e = \begin{cases} 0 & \text{if } c_l(d_0) - c_l(d_1) > c_r(d_0) - c_r(d_1) \\ 1 & \text{if } c_l(d_0) - c_l(d_1) < c_r(d_0) - c_r(d_1) \end{cases}$$

Output: extracted watermark chip $chip_s(k)$ and echo bit e

Experiment Results

We use Sound Quality Assessment Material (SQAM) audio [16] as test material, all 70 audio clips having a sampling frequency of 44100 Hz, 2 channels and a quantization of 16 bits. Various attacks are performed using Adobe Audition 3.0 and Audacity 1.3.6, which are both popular tool-sets for professional audio processing and editing.

Transparency Evaluation

SNR (Signal-to-Noise Ratio) [18] is a statistical difference metric which is used to measure the perceptual similarity between the undistorted original audio signal and the distorted watermarked audio signal.

Robustness Evaluation

Since an extracted watermark is taken as a proof of

authorship, the embedded watermark should withstand attempts at removing or damaging it. We have simulated five kinds of attempts, namely resampling, mp3 compression, white noise addition, low-pass filtering and high-pass filtering with audio processing software Adobe Audition 3.0 and Audacity, which can be freely downloaded from the Internet. The simulation results are listed below.

Another important criterion of watermarking algorithms is their minimum error rate of watermark detection or extraction. Therefore, we have used the CBR (Correct Bit Rate).

The correct bit rate in Table 7 are the averages of the correct bit rates obtained for 70 audio clips. The correct bit rate for each clip is defined as:

$$CBR = \frac{\text{Number of rightly extracted bits}}{\text{Number of embedded bits for the clip}}$$

TABLE 3 SNR OF ECHO HIDING AND ECHO HIDING WITH MCLT

audio sequence	SNR of our echo hiding method in Section 2.1	SNR of echo hiding with MCLT
clip 1-10	17.693	18.810
clip 11-20	18.784	22.810
clip 21-30	20.118	23.473
clip 31-40	22.147	15.650
clip 41-50	18.749	19.163
clip 51-60	17.249	19.358
clip 61-70	22.019	18.354
average	19.537	19.660

TABLE 4. CBR OF ECHO HIDING AND ECHO HIDING WITH MCLT UNDER VARIOUS ATTACKS

Attacks	Proposed echo hiding in Section 2.1	Our echo hiding with MCLT
No attacks	80.58%	94.71%
Re-sampling(22.05kHz)	82.83%	95.71%
MP3 compression	56.47%	86.30%
Noise attack	65.11%	77.86%
Low-pass filtering	73.37%	92.92%
High-pass filtering	64.15%	88.80%

Various options of the attacks above are defined as follows:

No attacks: closed loop(immediately decoding after encoding)

Re-sampling: sampling the watermarked signal with 22.05kHz sampling rate

MP3 compression: compressing the watermarked signal by MPEG-1 layer 3 and reverting it again to the original wave file

Noise attack: adding white noise with zero mean and Gaussian power density function to the watermarked

signal

Low-pass filtering: a first order low-pass filter with cut-off frequency 1600Hz is used

High-pass filtering: a first order high-pass filter with cut-off frequency 1600Hz is used

Table 3 and Table 4 show that combined with MCLT method, the robustness of echo hiding method is improved at the cost of lower embedding capacity and higher computational complexity.

According to some previous work [8], the correct bit rate of the echo method is no higher than 83%, which

is verified by our own tests. We take advantage of the majority vote to lower the bit error rate. We embed all the watermark bits 2^{k+1} times and then extract them. If no less than $k+1$ bits corresponding to one bit is 0, we decide that the bit is 0. If no less than $k+1$ bits corresponding to one bit is 1, we decide that the bit is 1. Our experimental results show that majority vote improves the bit correct rate and robustness of echo hiding method at the cost of lower capacity.

Conclusions

We first propose a new echo hiding method based on single echo hiding in [1], which is better than single echo hiding in both robustness and transparency with the same watermark capacity. Then, we have also improved the robustness of echo hiding method through combining it with the spread spectrum algorithm in MCLT domain [17]. Experimental results show that our two methods have high robustness.

ACKNOWLEDGMENTS

This research was supported in part by the National Natural Science Foundation of China (Grant Nos. 61070189, 61170032, and 61272435), Project of General Administration of press and publication of China under Grant GAPP-ZDKJ-BQ/15-2, and was also supported by the national postdoctoral foundation of China (Grand Nos. 201003111, 20110491838, and 2012T50842).

REFERENCES

- Bender W, Gruhl D, Morimoto N. Techniques for data hiding. IBM Systems Journal, 1996, vol.35, pp.313-336.
- Byeong-Seob Ko, Ryouichi Nishimura, Yoiti Suzuki. Time-spread echo method for digital audio watermarking. IEEE Transactions on Multimedia, 2005, vol.7(2), pp.212-221.
- Daniel Gruhl, Anthony Lu, Walter Bender. Echo hiding. IBM Systems Journal, 1996, vol.35, pp.328-332.
- Darko Kirovski, Henrique S.Malvar. Spread-spectrum watermarking of audio signals. IEEE Transactions on Signal Processing, 2003, vol.51 (4), pp.1020-1033.
- Foo Say Wei, Dong Qi. Audio watermarking of stereo signals based on echo-hiding method. Conference Proceedings of the 7th International Conference on Information, Communications and Signal Processing, 2009.
- H.S.Malvar. A modulated complex lapped transform and its applications to audio processing. IEEE International Conference on Acoustic, Speech, and Signal Processing, 1999, vol.3, pp.1421-1424.
- H.S.Malvar. Fast algorithm for the modulated complex lapped transform. IEEE Signal Processing Letters, 2003, vol.10, pp.8-10.
- Hyen O Oht, Jong Won Seoktt, Jin Woo Hongtt, Dae Hee Yount. New echo embedding technique for robust and imperceptible audio watermarking. International Conference on Acoustics, Speech, and Signal Processing, 2001, vol.3, pp.1341-1344.
- Hyoung Joong Kim, Yong Hee Choi. A novel echo-hiding scheme with backward and forward kernels. IEEE Transactions on Circuits and Systems for Video Technology, 2003, vol.13 (8), pp.885-889.
- Jingchang Chen, Ruimin Hu, Haojun Ai, Weiping Tu. DCT-II based fast algorithm for MCLT. Signal Processing, 2005, vol.21 (1), pp.63-66.
- Liuyun Duan, Jiebao Zong, Ying Wang, Hongfei Cheng, Zhong Yin. Developed cepstrum method for data extraction based on echo hiding. Electronic Design Engineering, 2010, vol.18 (7), pp.77-79.
- Qionghai Dai, Xinjian Chen. New algorithm for modulated complex lapped transform with symmetrical window function. IEEE Signal Processing Letters, 2004, vol.11 (12), pp.925-928.
- Quackenbush, S.R., Barnwell III, T.P., Clements, M.A.: 'Objective measures of speech quality' (Prentice-Hall, 1988)
- Seymour Shlien. The modulated lapped transform, its time-varying forms, and its Applications to audio coding standards, IEEE Transactions on Speech and Audio Processing, 1997, vol.5 (4), pp.359-366.
- SQAM-Sound Quality Assessment Material, <http://tech.ebu.cn/publications/sqamcd>
- Xinjian Chen, Qionghai Dai. A novel DCT-based algorithm for computing the modulated complex lapped transform. IEEE Transaction on Signal Processing, 2006, vol.54 (11), pp.4480-4484.
- Yousof Erfani, M. Shahram Moin, Mehdi Parviz. New methods for transparent and accurate echo hiding by

using the original audio cepstral content. 6th IEEE/ACIS International Conference on Computer and Information Science, 2007, pp.1087-1092.

Yun Ge, Dong Zhang. New Fast Algorithm for Modulated

Complex Lapped Transform with Arbitrary Windowing Function. Journal of Electronics & Information Technology, 2010, vol.32 (3), pp.747-749.